

УДК 330.13:623

DOI: [https://doi.org/1034169/2414-0651.2021.1\(29\).03-13](https://doi.org/1034169/2414-0651.2021.1(29).03-13)

І. І. ОЛІЙНИК, доктор технічних наук,
старший науковий співробітник

С. О. КОРОБЧЕНКО

<https://orcid.org/0000-0001-7650-5935>

(Центральний науково-дослідний інститут озброєння та військової техніки Збройних Сил України, м. Київ)

В. Г. ШОСТАК, в. о. директора Департаменту військово-технічної політики МОУ,
м. Київ, Україна

(Департамент військово-технічної політики МОУ, м. Київ)

МЕТОДОЛОГІЧНІ ОСНОВИ ПАРАМЕТРИЧНОЇ ОЦІНКИ ВАРТОСТІ ЖИТТЄВОГО ЦИКЛУ ОВТ

У статті розглянуто методологічні особливості застосування регресійного аналізу для проведення параметричної оцінки вартісних показників озброєння та військової техніки з використанням статистичних взаємозв'язків між історичними даними щодо витрат та тактико-технічними характеристиками складних технічних систем, проаналізовано процес параметричного оцінювання витрат, що складається з визначення основних оціночних гіпотез, збору даних щодо потенційних драйверів вартості, оцінки та нормалізації даних, проведення регресійного аналізу, тестування та відбору прийнятних моделей, а також досліджено основні обмеження та ризики щодо проведення таких оцінок.

Ключові слова: вартість життєвого циклу, параметрична оцінка, регресійний аналіз, мультиколінеарність, озброєння та військова техніка, NASA, метод найменших квадратів, інтервал прогнозування, скоригований коефіцієнт детермінації, множинна регресія.

ВСТУП

Наявність сучасних викликів державному суверенітету та територіальній цілісності України, нова спрямованість вектору її подальшого розвитку у бік Європейської та Євроатлантичної інтеграції зумовлюють необхідність внесення суттєвих змін в існуючу практику військового будівництва та систему оснащення ЗС України ОВТ сучасними зразками ОВТ.

У військово-технічній та воєнно-економічній сферах забезпечення воєнної безпеки держави мають бути реалізовані положення Закону України “Про оборонні закупівлі” [1], в якому у якості окремого критерію вводиться поняття “вартість життєвого циклу” озброєння, військової і спеціальної техніки та ставиться завдання щодо розробки НМА з його оцінки.

Використання розрахунку вартості ЖЦ повинно на кожному етапі програми підтримувати процес, за допомогою якого державні замовники мають можливість приймати оптимальні рішення за представленими їм варіантами. Ці варіанти можуть включати оцінку майбутніх видатків, порівняння альтернативних рішень, керування існуючими бюджетами і оцінку можливостей зниження витрат. Розрахунок вартості ЖЦ також використовується для оцінки повноти даних й визначення факторів витрат, пов'язаних із ключовими показниками ефективності використання системи або ключовими вимогами споживача [2].

Більшість оцінок вимагають використання різних методів розрахунку вартості ЖЦ [3, 4]. Однак, саме параметричний метод використовує статистично обґрунтовані предиктори та оцінює надійність припущень за допомогою використання реальних спостережень [18].

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

У зарубіжній та українській науковій літературі достатньої уваги приділено особливостям параметричної оцінки вартості. Так, питання існуючого інструментарію в галузі оцінювання витрат, що реалізований на основі відповідного науково-методичного апарату, розглянуті в роботах М. Хайхарста та Б. Вуда [5], П. Фріза, Д. Кловстада та ін. [6], М. Іванко, М. Домака та ін. [7], Р. Келого, Р. Бітена, Е. Махра [8], Л. Янга, Т. Джосеранда [9], Д. Хамакера [10], Е. Запати [11], Х. Джонса [12]. Крім того, питання параметричної оцінки проаналізовано в статтях А. Генрі [13], Д. Фішера [14], моделі оцінювання витрат досліджені в роботах К. Губера, К. Хореджсі, Р. Джежлевського [15], С. Ханта, М. Ванпелта [16] та І. Чепкова, В. Зубарева, В. Борохвостова, О. Рябця, І. Олійника та ін. [17].

Метою роботи є висвітлення підходів щодо застосування регресійного аналізу для проведення параметричної оцінки вартості ОВТ.

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

Параметричні оцінки витрат є результатом застосування методології оцінки витрат із використанням статистичних взаємозв'язків між історичними даними щодо витрат та іншими змінними, що стосуються складної технічної системи (наприклад, тактико-технічні та інші характеристики) для того, щоб розробити одну або кілька вартісних залежностей (CER).

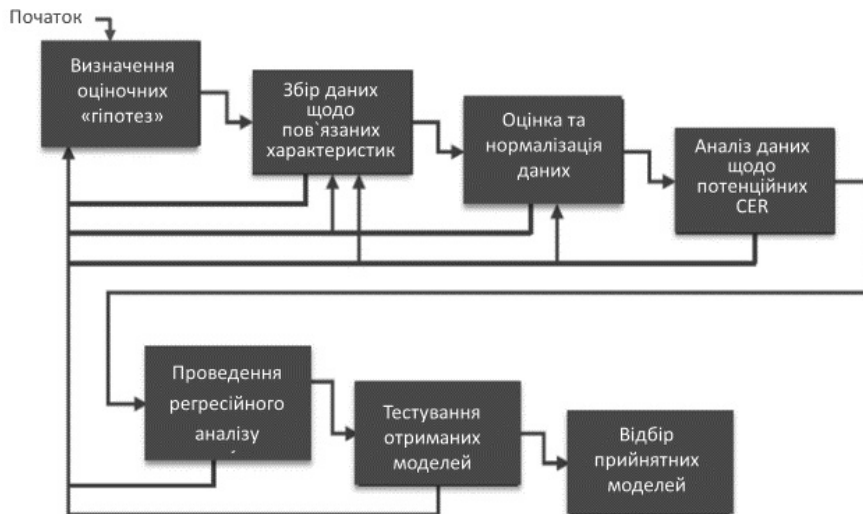


Рис. 1. Процес параметричної оцінки вартості

При застосуванні параметричної методології оцінювач має декілька альтернатив. По-перше, він може використати готові комерційні пропозиції вже розроблених CER; по-друге, він може застосувати загальновизнані рівняння або моделі. І в решті решт, він може створити власні CER.

Для розробки CER оцінювач повинен визначити фактори, які найбільше впливають на вартість.

Вивчивши технічну базу та проаналізувавши вхідні дані, оцінювач повинен перевірити вибрані драйвери витрат, обговоривши їх з інженерами, науковцями та / або іншими технічними експертами. Після цього CER може бути розроблена за допомогою математичного виразу, який може варіюватися від простого емпіричного правила до рівняння, що використовує багато параметрів (характеристик) у якості незалежних змінних.

Оцінки, створені з використанням параметричного підходу, базуються на історичних даних та математичних виразах, що використовують вартість як залежну змінну до обраних, незалежних змінних, що впливають на витрати.

Регресійний аналіз є основним методом, за допомогою якого здійснюється параметричне оцінювання витрат. Цей метод відноситься до прикладної статистики та застосовується для кількісного визначення (математичного виразу) зв'язку між змінними, а також подальшого опису точності цього співвідношення за різними показниками. В більшості випадків параметричної оцінки вартості для оцінки невідомих параметрів регресійних моделей за вибірковими даними використовується метод найменших квадратів (МНК).

Основне припущення про параметричну оцінку передбачає, що ті самі фактори, які впливали на вартість у минулому, впливатимуть на вартість у майбутньому. Наприклад, у разі оцінки витрат життєвого циклу космічних систем в NASA [3] у якості так званих драйверів вартості часто використовують вагові характеристики та складність конструкції літальних апаратів.

На рис. 1 показано кроки, пов'язані з параметричним оцінюванням витрат.

Основною проблемою при аналізі біваріантних (дві змінні) та багатоваріантних (три або більше змінних) даних є виявлення та вимір зв'язку чи коваріації між змінними – тобто визначення ступеню взаємозалежності змінних. Коли взаємозв'язок між змінними є чітким і точним, звичайних математичних методів вистачає. Коли відносини розмиті або неточні, перевагу надають використанню саме статистичних методів.

Ми можемо оцінити, наскільки значною є невизначеність та чи є взаємозалежність взагалі. Якщо є лише помірна кількість невизначеності, можна підрахувати, яким є найкращий прогноз, а також оцінити його якість, враховуючи неточність взаємозв'язків. Є два пов'язані, але різні аспекти вивчення зв'язку між змінними.

Перший, регресійний аналіз, який передбачає спробу встановити природу взаємозв'язку між змінними – тобто вивчити функціональний взаємозв'язок між змінними і тим самим забезпечити механізм прогнозування.

Другий, кореляційний аналіз, що має на меті визначити ступінь взаємозв'язку між змінними.

Якщо існує зв'язок між будь-якими змінними, існує чотири основні причини для цього.

1. Першою причиною є випадок. Це загальновідомий тип несподіваних і незрозумілих подій. В такому варіанті тип взаємозв'язку між змінними не піддається кількісній оцінці.

2. Другою причиною взаємозв'язку між змінними може бути зв'язок з третьою сукупністю обставин (впливаючих факторів). В цьому випадку відносини піддаються кількісній оцінці, але потребують визначення прямих зв'язків.

3. Третя причина для кореляції – це зрозуміла функціональна залежність, яку можна представити у вигляді рівнянь. Такий тип точної залежності рідко існує при оцінці витрат.

4. Остання причина – причинно-наслідкові зв'язки. Ці співвідношення також представлені рівняннями, але в цьому випадку між змінними виводяться причинно-наслідкові залежності. Слід зазначити, що регресійний аналіз не доводить причини та наслідки. Натомість регресійний аналіз представляє аналітикові можливості припустити, що встановлений зв'язок між змінними є логічним та припустимим.

Отже, в регресійному аналізі існує два різні типи змінних: залежні змінні, що позначаються символом Y та незалежні змінні, що позначаються символом X . При цьому залежна змінна реагує на зміни незалежної змінної. У випадку пошуку прийнятних CER змінна Y представляє вартість, тоді як змінна X представляє різні параметри системи. Як зазначено вище, регресійний аналіз використовується не для підтвердження причинності, а для представлення висновку про причинність. Іншими словами, незалежно від статистичної значущості результату регресії, причинність не може бути доведена.

Наприклад, припустимо, що в рамках проекту з розроблення силової установки бойового літака оцінювач хоче спрогнозувати її вартість, виходячи з заданих тактико-технічних вимог. Оцінювач досліджує, наскільки певні характеристики корелюють по відношенню до вартості. Якщо певні системні вимоги (наприклад, тяга) вказують на сильну кореляцію із вартістю системи, і ці зв'язки виглядають логічними (наприклад має місце позитивна кореляція), тоді можна зробити висновок, що ці рівняння мають причинно-наслідковий зв'язок.

Хоча регресійний аналіз не може підтвердити причинність, він надає інструмент для вимірювання сили взаємозв'язків у кількісному вигляді та оцінки і перевірки гіпотез стосовно параметрів моделі. Перед проведенням регресійного аналізу важливо вивчити та нормалізувати дані. Цей процес може включати наступні заходи:

- приведення вартості до базового року з урахуванням коефіцієнту інфляції;
- коригування кривої навчання до загальної заданої одиниці (наприклад, вартості першої одиниці);
- перевірка незалежних змінних із застосуванням екстраполяції;
- проведення аналізу діаграми розсіювання;
- перевірка однорідності даних;
- перевірка незалежних змінних на наявність мультиколінеарності, тощо.

Першим кроком при проведенні регресійного аналізу є пошук незалежної змінної або змінних (наприклад, вага системи, X), що можуть мати значний вплив на залежну змінну (наприклад, вартість системи, Y).

Цей крок зазвичай виконується шляхом створення діаграми розсіювання пар даних (X , Y), після чого можливо визначити певну тенденцію. Для CER залежна змінна завжди буде представляти вартість, а кожна незалежна змінна в свою чергу буде драйвером витрат. Кожен драйвер витрат слід вибирати лише тоді, коли між ним та вартістю існує значна кореляція та зрозуміла природа існуючих взаємозв'язків. Наприклад, аналітики припускають, що складність (X) комп'ютерного програмного забезпечення обумовлює його вартість (Y) та можуть дослідити свої припущення шляхом побудови пар цих залежних (Y) та незалежних (X) змінних. Приклад побудови діаграми розсіювання статистичних даних про вартість (Y) по відношенню до ваги системи (X) наведено на рис. 2.

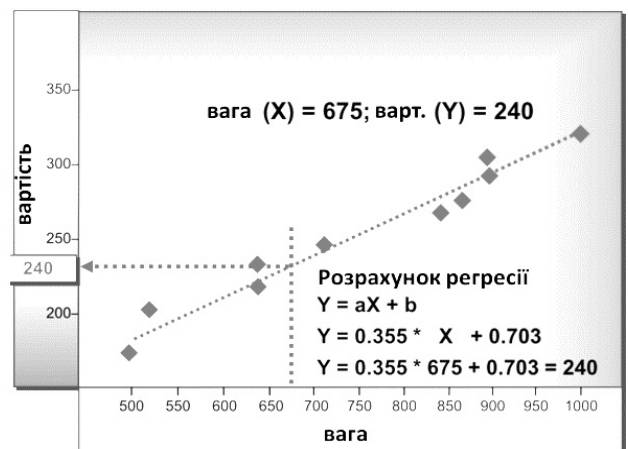


Рис. 2. Діаграма розсіювання та лінія регресії даних про вартість (Y) по відношенню до ваги (X)

Завдання регресійного аналізу полягає в тому, щоб «підігнати» лінію регресії якомога точніше до даних та отримати рівняння, яке описує цю лінію, наступного виду:

$$Y = Y(\text{перетин}) + (\text{нахил}) \cdot (X). \quad (1)$$

За результатом аналізу на рис. 2 можна припустити позитивну кореляцію, так як при збільшенні ваги зростає і вартість. Набагато рідше зустрічається CER з негативною кореляцією.

Наступним кроком у проведенні регресійного аналізу для отримання CER є обчислення зв'язку між залежною (Y) та незалежною (X) змінними. Іншими словами, проводиться статистична перевірка обґрунтованого ступеню причинності і наслідковості. Показники, як правило, підпорядковуються лінійному або нелінійному закону розподілення. Для лінії регресії на рис. 2 умовний CER зображений як лінійна залежність вартості: Вартість = $A + B \cdot \text{Вага}$.

Як зазначалося раніше, найпоширенішим методом побудови регресії є МНК, який можна застосовувати для моделювання як лінійних, так і нелінійних CER. Ключові позначення, що використовуються в регресійному аналізі, наведені в табл. 1.

Таблиця 1

Ключові символи, що використовуються в регресійному аналізі

Позначення	Опис	Визначення	Оцінка
X, Y	Дані спостережень	Y – залежна змінна X – незалежна змінна	Мають бути перевірені та виправлені будь-які помилки, особливо відхилення в даних. Якщо якість даних є неприйнятною, може бути застосований інший метод аналізу ніж регресійний аналіз
$\bar{X}, \bar{Y}$	Середні значення	$\bar{X}$ – середнє значення X_i $\bar{Y}$ – середнє значення Y_i	Допомагають описати основні тренди даних, що оцінюються
$\hat{Y}_i$	Розраховане Y	$\hat{Y}_i$ – залежна змінна	Розраховується за допомогою МНК та представляє собою спрогнозоване або “підлаштоване” значення, що залежить від X_i та відрізняється від фактичних значень Y_i
$\hat{b}_i, \hat{B}_i$	Розрахований регресійний параметр (коефіцієнт) для кожної незалежної змінної	Значення «перетину» та кожного «нахилу» у рівнянні регресії	Якщо t-статистика нижче порогового значення або значення не відповідають логічному критерію, регресійна модель має бути перероблена
e_i	Залишки (помилки)	Різниця між фактичними та розрахованими значеннями Y	Мають бути перевірені та за необхідності виправлені помилки трактування
R^2	Коефіцієнт детермінації	Вимірює ступінь відповідності моделі фактичним даним	Чим ближчим є значення R^2 до 100 %, тим кращою є відповідність
R_a^2	Скоригований коефіцієнт детермінації	R^2 , що враховує вплив однієї або декількох пояснюючих змінних	Одним з доказів того, що незалежна змінна, що додається, немає значної описової здатності, є зниження скоригованого коефіцієнту детермінації
SST	Сума квадратичного сукупного	$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2$	Використовується для розрахунку R^2 та R_a^2 . Чим вищим є значення SST , тим більшою є дисперсія фактичних значень Y від їх середнього значення
SSR	Сума квадратичної регресії (пояснене відхилення)	$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	Використовується для розрахунку R^2 та R_a^2 . Чим вищим є значення SSR , тим більше лінія регресії відрізняється від середнього значення Y
SSE	Сума квадратичної помилки	$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	Використовується для розрахунку R^2 та R_a^2 . Чим вищим є значення SSE , тим більшою є дисперсія спрогнозованих від фактичних значень Y

Модель простої лінійної регресії

«Проста регресія» означає той факт, що для прогнозування значення Y потрібна лише одна незалежна змінна. В більшості випадків під час розроблення CER використовується МНК. Хоча це може здаватися симпліфікацією завдання, існує кілька вагомих причин для використання такого підходу. Витрати дуже часто змінюються лінійно по відношенню до фізичних та експлуатаційних характеристик системи. Навіть якщо такі залежності не є прецизійно лінійними, їх лінійна

апроксимація є досить прийнятною, особливо враховуючи точність вхідних даних. Крім того, більшість криволінійних та експоненційних функцій можуть бути перетворені в лінійну форму, тим самим піддаючись лінійному аналізу. І нарешті, наш обсяг вибірки часто настільки малий, що не може виправдати використання будь-якої іншої форми.

У моделі, що заснована на МНК, залежна або пояснювана змінна Y відноситься до незалежної або пояснювальної змінної X за таким виразом:

$$Y = \beta_0 + \beta_1 X + E. \quad (2)$$

Цей вираз є моделлю МНК для сукупності, де β_0 , Y та β_1 є невідомими параметрами регресії, що називаються коефіцієнтами регресії, а E – випадкова помилка або залишок.

Позначення змінної як залежної чи незалежної стоюється її математичної або функціональної залежності та не передбачає ні статистичної залежності, ні причинно-наслідкового зв'язку.

Слід зазначити, що згадана раніше модель МНК має дві окремі частини: систематичну частину і стохастичну або випадкову (E). Цей розподіл показує, що модель скоріше імовірна, ніж детермінована. Стохастичний характер регресійної моделі передбачає, що значення Y ніколи не може бути прогнозовано так само, як у детермінованому випадку. Невизначеність щодо Y пояснюється наявністю E .

Для того, щоб забезпечити обґрунтованість використання МНК, слід зробити п'ять припущень.

1. Функціональна форма вказана правильно. Це означає, що між X і Y існує лінійна залежність, і що достатньо лише однієї незалежної змінної, щоб пояснити варіацію залежної змінної.

2. Вважається, що незалежні змінні вимірюються без помилок.

3. Залишки E вважаються нормально розподіленими навколо лінії регресії. Це припущення дозволяє використовувати певні статистичні дані та тести на значимість для оцінки валідності лінії регресії. Відомо, що різниця між двома випадковими величинами, незалежно від їх початкового розподілу, як правило, поширюється нормально.

4. Залишки E приймаються з ідентично і незалежно розподіленого випадкового розподілу із нульовим математичним очікуванням та постійною дисперсією. Це означає, що залишки не можуть бути спрогнозовані на основі даних про незалежну змінну X .

5. Вихідні дані є однорідними.

Використовуючи ці припущення, можемо отримати оцінки для невідомих параметрів регресії β_0 та β_1 та зробити необхідні висновки. Слід підкреслити, що на практиці одне або кілька з припущень часто порушуються. Часто незалежні змінні не вимірюються без помилок, а обсяг вибірки настільки малий, що припущення про нормальність є невірними. Однак істотною частиною регресійного аналізу є перевірка цих припущень (або перевірка їх порушення). Для цього використовують методи (інструменти) аналізу залишків.

Регресійна модель МНК для сукупності представляє рівняння прямої лінії. При цьому еталонна модель буде мати наступний вигляд:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + E. \quad (3)$$

В цьому рівнянні позначення « $\wedge$ » вказує на спрогнозоване (тобто обчислене) значення Y та обчислені значення нахилу ($\hat{\beta}_0$) та перетину ($\hat{\beta}_1$) лінії регресії.

Подібно моделі МНК для сукупності, X є незалежною змінною.

Розрахунок коефіцієнтів для моделі простої лінійної регресії

Наступні два рівняння дозволяють розрахувати значення нахилу і перетину лінії регресії:

$$\hat{B}_1 = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}; \quad \hat{b}_0 = \frac{\sum Y - \hat{b}_1 \sum X}{n}. \quad (4)$$

Вирішуючи наведені вище рівняння, ми повністю визначаємо лінію регресії. Тому для будь-якого значення незалежної змінної (значення певного параметру/характеристики системи), в межах заданих значень незалежної змінної, ми можемо визначити спрогнозоване «середнє» значення залежної змінної (вартості).

Екстраполяція поза відповідним діапазоном CER

Прогнози з використанням CER є дійсними лише тоді, коли застосовуються у відповідному діапазоні значень незалежної змінної, що містяться в наборі даних. При цьому екстраполяція поза діапазоном даних є небезпечною, оскільки не можна зробити висновок щодо поведінки даних за межами діапазону вибірки. Крім того, екстраполяція може використовуватися при регресії даних динамічних рядів, оскільки передбачається, що ця залежність триває у часі.

Статистика регресії простих лінійних моделей

Враховуючи те, що кількісне визначення відносин між незалежними та залежними змінними становить лише половину проблеми регресійного аналізу, необхідно обчислити регресійну статистику та провести тести значимості, які вимірюють валідність лінії регресії.

При цьому регресійна статистика допомагає нам визначити прогностичну здатність CER. Вона також може відповісти на запитання, чи існує певний тренд. Щоб вирішити, чи використовувати рівняння регресії, оцінювачу слід принаймні розрахувати, а потім оцінити коефіцієнт детермінації (R^2), стандартну похибку оцінки (SE), коефіцієнт варіації (CV) та «силу» коефіцієнтів регресії (перевірка гіпотез або «t-тести»).

Розрахунок коефіцієнта детермінації (R^2) для рівняння простої лінійної регресії (SLR)

Коефіцієнт детермінації є одним з найбільш застосовуваних типів регресійної статистики, що приймає значення від 0 до 1 (або 0 відсотків і 100 відсотків). Чим більше R^2 , тим краще «відповідає» лінія регресії фактичним даним. Наприклад, у випадку простої лінійної регресії $R^2=0,93$ означає, що 93 відсотки варіації вартості можна пояснити варіацією незалежної змінної.

Для того, щоб розрахувати R^2 , спочатку потрібно зрозуміти наступні терміни: пояснене відхилення, незрозуміле відхилення та повне відхилення. Розглянемо наведену на рис. 3 лінію регресії

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X.$$

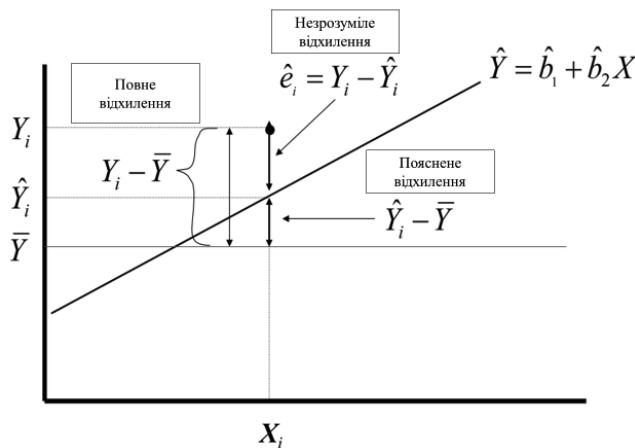


Рис. 3. Ілюстрація відхилень від середнього значення та лінії регресії

Як видно з графіку, окреме відхилення фактичного значення Y_i від середнього значення $\bar{Y}$ можна розділити на дві частини: незрозуміле та пояснене відхилення.

При цьому:

1. X_i представляє i -те значення незалежної змінної;
2. Y_i – фактичне спостережуване значення вартої, пов'язане з X_i ;
3. $\hat{Y}_i$ – значення залежної змінної, отримане за допомогою МНК (спрогнозоване значення);
4. $\bar{X}$ та $\bar{Y}$ є середніми значеннями фактичних значень X_i та Y_i ;
5. $\bar{Y}$ завжди має перетин на лінії регресії;
6. $\bar{Y}$ використовується у якості орієнтиру, від якого потрібно виміряти кожне «пояснене відхилення» та «повне відхилення» та є базисом для коефіцієнта детермінації.

Для того, щоб розрахувати коефіцієнт детермінації, кожен з трьох зазначених типів відхилення має бути зведений у квадрат, а потім підсумований. Таким чином, отримаємо наступне:

1. Пояснене відхилення дає суму квадратичної регресії (SSR), де

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2. \quad (5)$$

2. Незрозуміле відхилення дає суму квадратичної помилки (SSE), де

$$SSE = \sum (Y_i - \hat{Y}_i)^2. \quad (6)$$

3. Повне відхилення утворює суму квадратичного сукупного (SST), де

$$SST = \sum (Y_i - \bar{Y})^2. \quad (7)$$

Оскільки SSR або R^2 є відношенням частини до цілого, воно матиме значення від 0 до 1. У разі використання для оцінки CER R^2 має бути більше або дорівнювати **90 відсоткам (0,9)**. При цьому коефіцієнт детермінації може використовуватися для порівняння різних ліній

регресії для визначення найкращої лінії пристосування (лінія регресії з найвищим R^2).

Цей статистичний показник по суті вимірює відношення пояснюваного відхилення до загального відхилення. Так як пояснене відхилення представляє здатність лінії регресії прогнозувати, вона повинна бути якомога більшою часткою повного відхилення. Іншими словами, враховуючи, що SST є постійним, чим вищий SSE, тим вищий є значення R^2 .

$$R^2 = \frac{SSR}{SST} = \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (\hat{Y}_i - \bar{Y})^2}. \quad (8)$$

З формули (8) видно, що лінія регресії, яка найкраще відповідає точкам вибірки, має місце, коли сума квадратів незрозумілих відхилень є мінімальною. Кожне незрозуміле відхилення – це вертикальна відстань від лінії до фактичної точки даних. Саме цей критерій робить лінію регресії якомога ближче до всіх точок. Іншими словами, враховуючи, що SST є постійним, чим менший SSE, тим вище значення R^2 .

Розрахунок стандартної помилки оцінки (SE)

Стандартна похибка оцінки аналогічна стандартному відхиленню вибірки. Це міра відхилення точок вибірки від лінії регресії. Хоча R та R^2 також вимірюють якість «відповідності» лінії регресії до точок даних, вони є лише відносними мірами і на них впливає нахил лінії регресії. Стандартна помилка оцінки є абсолютною мірою відхилення, і її знак не впливає на нахил лінії регресії. Стандартна помилка оцінки є квадратним коренем із суми квадратів незрозумілого відхилення, що поділена на ступені свободи, і позначається як SE.

$$SE = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-k-1}} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{MSE}, \quad (9)$$

де n – розмір вибірки;

k – кількість незалежних змінних;

MSE – середньоквадратична похибка.

Оскільки $\hat{b}_0$ та $\hat{b}_1$ визначаються з тих самих точок даних, з яких обчислюється SE, має бути врахована відповідна кількість ступенів свободи (як мінімум, $n - 2$). При цьому позначення k відноситься до кількості незалежних змінних, що містяться в CER. Для простих лінійних рівнянь, де існує лише одна незалежна змінна, має бути розраховане тільки одне значення нахилу.

Чисельник рівняння SE ($\sum (Y_i - \hat{Y}_i)^2$) називається сумою квадратичних (незрозумілих) помилок, або SSE. SSE, поділений на ступінь свободи, називається середньою квадратичною помилкою (MSE), яка є дисперсією помилок.

Як вже зазначалося, SE є абсолютною мірою відхилення. Отже, вона матиме такі самі одиниці вимірювання, як і залежна змінна. Єдиним обмеженням для SE є те, що її значення не може бути негативним. Через це стандартна помилка оцінки не може бути використана для самостійної та повної оцінки лінії регресії, а лише для

порівняння різних ліній регресії. Стандартна помилка оцінки також використовується для обчислення інших статистичних даних та тестів значущості. Оскільки SE є мірою незрозумілої помилки, вона повинна бути якомога меншою. Отже, формула для розрахунку SE може бути записана у наступному вигляді:

$$SE = \sqrt{\frac{\sum Y^2 - \hat{b}_0 \sum Y - \hat{b}_1 \sum XY}{n-2}} \quad (10)$$

Довірчий інтервал та інтервал прогнозування моделей простої лінійної регресії

Довірчий інтервал та інтервал прогнозування представляють собою інтервали на лінії регресії, які надають оцінювачу додаткову інформацію щодо отриманої CER. Довірчий інтервал представляє область, в якій з імовірністю $(1 - \alpha)$ буде лежати лінія регресії. Цей інтервал фактично дає граничні межі для «нахилу» лінії. Водночас, інтервал прогнозування вказує на діапазон, на який передбачається попадання певного відсотка спрогнозованих значень залежної змінної.

Завдяки цій особливості, зазначені інтервали можна використовувати для передбачення та оцінки майбутніх (гіпотетичних) значень вартості в межах заданого діапазону та заданого рівня впевненості.

Інтервал прогнозування існує завдяки відхиленню між фактичними (Y_i) та розрахованими ($\hat{Y}_i$) значеннями залежної змінної. Стандартне відхилення цього розподілу становить:

$$SE \sqrt{\frac{n+1}{n} + \frac{(X_0 - \bar{X})^2}{\sum X^2 - n\bar{X}^2}} \quad (11)$$

де X_0 – значення X , для якого ми обчислюємо інтервал прогнозування.

З урахуванням цієї інформації, послідовність розрахунку інтервалу прогнозування може бути наступною.

1. Визначити рівень значущості (α).
2. Знайти $t_p = t_{(1-\alpha/2, n-2)}$ з t -таблиці.
3. Визначити стандартну похибку оцінки, SE.
4. Вибрати значення X_0 для оцінки.
5. Побудувати інтервал прогнозування.

$$PI = \hat{Y}_0 \pm t_p SE \sqrt{\frac{n+1}{n} + \frac{(X_0 - \bar{X})^2}{\sum X^2 - n\bar{X}^2}} \quad (12)$$

На рис. 4 наведений приклад інтервалу прогнозування для параметричної оцінки вартості системного елемента в залежності від його ваги.

Хоча інтервал прогнозування дуже схожий на довірчий інтервал, він завжди буде трохи ширшим. Зазвичай, інтервали прогнозування розраховують для конкретної точки, що лежить на осі Y , а не для лінії регресії в цілому.

Моделі множинної регресії

Існує багато ситуацій, коли для того, щоб адекватно описати вартісні залежності та зробити коректні висновки, потрібно використовувати дві або більше незалежних змінних. При цьому, регресійні моделі,

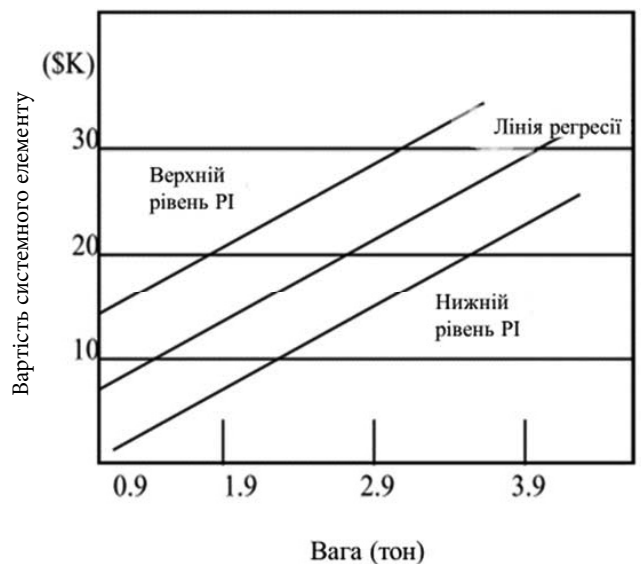


Рис. 4. Ілюстрація інтервалу прогнозування для лінійної регресії вартості системного елемента в залежності від його ваги

що використовуються у таких випадках, називають моделями множинної регресії. Прикладом такої моделі може бути прогнозування потреби у запчастинах для певної системи озброєння, в якій в якості незалежних змінних застосовують середній час роботи, середній пробіг та середньодобову температуру або прогнозування вартості зразка ОБТ в залежності від його тактико-технічних характеристик.

В загальному вигляді модель множинної регресії може бути представлена таким виразом:

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \dots + \hat{b}_k X_k \quad (13)$$

де $\hat{Y}$ – розраховане значення залежної змінної;

X_i – i -та незалежна змінна;

k – кількість незалежних змінних;

$\hat{b}_i$ – i -й параметр регресії (коефіцієнт незалежної змінної).

Так, зважаючи на велику кількість ТТХ (як правило, більше 10), у випадку прогнозування вартості зразка ОБТ постає питання відбору найбільш значимих для моделі вхідних ознак. Скорочення кількості незалежних змінних має на меті зменшення розмірності моделі та покращення її якості шляхом усунення проблеми надмірності та мультиколінеарності.

Вочевидь, що одним з можливих виходів з ситуації могла б стати побудова моделі на всіх можливих комбінаціях наборів вхідних ознак з подальшим відбором того варіанту, який мав би найкращу описову здатність результуючої ознаки і при цьому містив би оптимальну кількість незалежних змінних. Однак таке рішення можливо лише при наявності незначної кількості чинників-претендентів на включення в модель. У разі ж наявності великого списку потенційних ознак подібна методика представляється досить складною, так як кількість моделей, які необхідно буде побудувати, вияв-

ляється вкрай великою і в загальному випадку дорівнює $2n-1$ (так зване «прокляття» розмірності).

Зважаючи на це, для вирішення цієї проблеми, доцільно використовувати методи добору незалежних змінних що ґрунтується на використанні приватного F-критерію (Фішера), наприклад процедуру прямого добору (Forward Selection).

Для перевірки валідності зазначеної моделі має бути перевірено наявність декількох припущень, що вже згадувались для випадку моделі простої лінійної регресії:

1. Функціональна форма залежності визначена правильно.
2. Значення незалежних змінних наведені без помилок.
3. Залишки ($e_i = Y_i - \hat{Y}_i$) є незалежними, нормально розподіленими, з нульовим середнім значенням та постійною дисперсією та не мають взаємної кореляції.
4. Вихідні данні є однорідними.
5. Незалежні змінні не мають взаємної кореляції.

У разі множинної регресії значення параметрів розраховуються також за допомогою МНК в рамках рішення системи нормальних рівнянь.

Скоригований коефіцієнт детермінації R^2_{adj}

Коефіцієнт детермінації для моделей множинної регресії (множинний коефіцієнт детермінації) визначається аналогічно коефіцієнту детермінації простих лінійних моделей. Він вимірює пропорційне зменшення повного відхилення від середнього значення $\bar{Y}$, пов'язаного з усіма незалежними змінними в моделі множинної регресії.

$$R^2 = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2} \quad (14)$$

Коефіцієнт множинної детермінації може бути модифікований, щоб врахувати кількість незалежних змінних у моделі. Ця модифікація коригує втрату ступенів свободи внаслідок додавання незалежних змінних. По суті, R^2 може або збільшуватися, або залишатися без змін при додаванні додаткових незалежних змінних. Це обумовлюється тим, що пояснене відхилення може зростати, тоді як загальне відхилення завжди залишається однаковим. Модифікований коефіцієнт враховує втрату ступеня свободи для кожної додаткової незалежної змінної. Такий показник називається скоригованим коефіцієнтом детермінації R^2_{adj} та визначається за формулою:

$$R^2_{adj} = 1 - \frac{(1-R^2)(n-1)}{(n-k-1)}, \quad (15)$$

де k – кількість незалежних змінних x_i ;

n – кількість спостережень.

R^2_{adj} може зменшуватися, коли в модель вводиться інша незалежна змінна, втрата ступеню свободи в знаменнику $(n-k-1)$ може з лишком компенсувати зменшення незрозумілих відхилень. Якщо всі CER мають однаковий тип (прості лінійні, множинні) та однакову кількість незалежних змінних, обчислювати R^2_{adj} не потрібно.

Особливості розрахунку стандартної помилки оцінки (SE) для множинних моделей регресії

Стандартна помилка оцінки (SE) для множинних моделей регресії розраховується як і для випадку простої лінійної моделі. При цьому з формули (10) видно, що кількість незалежних змінних безпосередньо впливає на стандартну похибку оцінки.

Так, за рівних інших умов, при зростанні кількості незалежних змінних, зростає і SE. Це відбувається тому, що збільшення « k » веде до зменшення значення знаменника « $n-k-1$ ».

Для того, щоб формула SE мала рішення в діапазоні «дійсних чисел», обсяг вибірки (n) повинен перевищувати кількість незалежних змінних (k), принаймні на одну.

Формулювання висновків про коефіцієнти регресії для моделей множинної регресії (гіпотетичні тести)

На відміну від простої лінійної моделі у разі множинної регресії перевіряються не тільки окремі нахили, але й загальне рівняння. При цьому для встановлення факту, чи є відношення регресії між залежною змінною та набором незалежних змінних статистично значущим, може використовуватися F-критерій (Фішера), критерій Кохрена або ін.

Основними етапами виконання перевірки гіпотези щодо значимості є:

1. Встановити нульову та альтернативну гіпотези.
 - а) Нульова гіпотеза. $H_0 : b_1 = b_2 = \dots = b_k = 0$.
 - б) Альтернативна гіпотеза. $H_1 : \text{не всі } b_k = 0$.
2. Визначити рівень значущості (α), необхідний для тесту.
3. Знайти $F_{p(1-\alpha, k, n-k-1)}$ з таблиці F-розподілення.
4. Порахувати $F_c = \frac{R^2/k}{(1-R^2)/(n-k-1)}$.
5. Прийняти рішення на основі наступних правил.
 - а) Якщо $F_c > F_p$ – відхилити нульову гіпотезу та зробити висновок, що відношення регресії між залежною змінною та набором незалежних змінних є статистично значущим.
 - б) Якщо $F_c \leq F_p$ – прийняти нульову гіпотезу та зробити висновок, що відношення регресії між залежною змінною та набором незалежних змінних **не** є статистично значущим та відповідно слід вважати дану модель непринятною.

Проблема мультиколінеарності в моделях множинної регресії

У разі існування взаємних залежностей між незалежними змінними (це може бути одне або кілька значень X_i , виражене через інші значення X_k або лінійну комбінацію цих значень), МНК не може бути застосований в загальному вигляді. Така ситуація, при якій сукупність спостережень різних незалежних змінних мають високу взаємну кореляцію, називається мультиколінеарністю. У цьому випадку, ми стикаємось з двома негативними наслідками:

розрахункове значення коефіцієнта регресії b_k (для даної незалежної змінної), може сильно варіюватися залежно від того, які інші незалежні змінні є в моделі.

Отже, значення коефіцієнта насправді не вказує на внесок цієї конкретної змінної;

коефіцієнти регресії, як правило, мають надзвичайно великі стандартні відхилення.

Для усунення мультиколінеарності можуть бути вжиті наступні заходи:

отримання додаткових спостережень для того, щоб зламати шаблон мультиколінеарності;

перетворення деяких незалежних змінних для того, щоб зменшити ступінь мультиколінеарності;

зміна моделі шляхом видалення деяких незалежних змінних, що спричиняють мультиколінеарність.

Для перевірки мультиколінеарності може бути використана матриця кореляції, що надається більшістю спеціалізованих програмних статистичних пакетів або проведений розрахунок лінійної регресії між усіма парами незалежних змінних.

Для будь-якої пари незалежних змінних, якщо $R^2 > 0.5$, ($|R| > 0.7$), то існує залежність між незалежними змінними, яка вірогідно буде впливати на параметри регресії. При цьому, всі результати тестів на значущість та інші висновки будуть недійсними [18]. Цю пару незалежних змінних не слід використовувати разом у регресійній моделі. Для перевірки мультиколінеарності для кожної пари незалежних змінних може бути використана наступна формула:

$$R_{X_1, X_2} = \frac{\sum X_1 X_2 - n \bar{X}_1 \bar{X}_2}{\sqrt{\sum X_1^2 - n \bar{X}_1^2} \sqrt{\sum X_2^2 - n \bar{X}_2^2}}. \quad (16)$$

ВИСНОВКИ

Отже, хоча регресійний аналіз в контексті побудови вартісних залежностей і є дуже корисним інструментом, він часто несе багато ризиків щодо неправильного тлумачення отриманих результатів. Причому найбільшу загрозу щодо неправильної інтерпретації несе плутанина між статистичною залежністю та причинно-наслідковим зв'язком.

Взаємовідносини між статистичною залежністю та причинно-наслідковим зв'язком можуть бути сформульовані наступним чином: наявність статистичної залежності не обов'язково передбачає наявність причинно-наслідкового зв'язку, але наявність останнього завжди передбачає статистичну залежність (тобто кореляцію). При цьому статистичні докази можуть показати лише ступінь асоціації між змінними. Судження щодо наявності причинно-наслідкового зв'язку залежить суто від міркувань (логіки).

Крім того, хибні висновки також можуть бути результатом екстраполяції поза діапазоном даних. Регресійний аналіз – це спроба встановити стосунки в межах діапазону. Відповідно відносини в різних діапазонах можуть кардинально відрізнятися. До невірних результатів також можуть вести помилки у вихідних даних та наявність сторонніх спостережень, що обумовлюють середнє значення, оскільки саме від цього значення мінімізується відхилення в рамках регресійного аналізу.

Тому, враховуючи вищезазначені ризики, процес відбору найбільш прийнятної моделі може включати на-

ступні кроки, послідовність виконання яких може дещо відрізнятися.

1. Обрання моделі із максимальним R^2_{adj} або мінімальним SE.

2. Проведення тестів на значимість.

3. Перевірка моделі на наявність мультиколінеарності.

4. Проведення аналізу залишків.

5. Проведення логічного тесту, (підтвердження того, що відносини, описані в CER, мають сенс).

У разі невідповідності моделі на кроках (2–5) необхідно перейти до кроку 1 та обрати наступну модель з максимальним R^2_{adj} . Нарешті, у випадку відсутності прийнятної моделі CER мають бути застосовані непараметричні методи визначення вартості.

Ураховуючи викладені у статті особливості застосування регресійного аналізу для параметричної оцінки вартості життєвого циклу ОБТ, доцільно врахувати їх під час розроблення науково-методичного апарату з оцінки вартості ЖЦ ОБТ, який дозволить забезпечити замовника потрібною інформацією про вартісні показники зразків ОБТ для прийняття обґрунтованих рішень стосовно порівняння альтернатив, фінансового планування, контролю витрат тощо.

При цьому, на думку авторів, подальші дослідження у цьому напрямку слід зосередити на розробленні конкретних алгоритмів, направлених на побудову математичних моделей, необхідних для прогнозування.

СПИСОК ПОСИЛАНЬ

1. Закон України «Про оборонні закупівлі № 808-IX від 17 липня 2020 року». Відомості Верховної Ради України. 2020. № 67. Ст. 223.
2. NATO Publication ALCCP-01. NATO GUIDANCE ON LIFE CYCLE COSTS.
3. Коробченко С.О. Процес оцінки витрат за проектами в NASA. Озброєння і військова техніка. Київ: ЦНДІ ОБТ ЗС України. 2019. № 4. С. 3—11.
4. Коробченко С.О., Шаповал П.І. Аналіз системи оцінки вартості життєвого циклу зразків озброєння та військової техніки в країнах – членах НАТО. Зб. наук. пр. Київ: ЦНДІ ОБТ ЗС України. 2020. № 3(78). С. 224—236.
5. Hayhurst, M.R. & Wood, B.W. (2017). Mission Operations Cost Estimation Tool (MOCET). 2017 IEEE. Available at: <https://ieeexplore.ieee.org/abstract/document/7943686>.
6. Friz, P.D., Klovstad, J.J., Leser, B.B. & Towle, B.C. (2018). Blind Study Validating Parametric Costing Tools PRICE True Planning and SEER-H for NASA Science Missions. AIAA SPACE. Available at: <https://arc.aiaa.org/doi/abs/10.2514/6.2018-5178>.
7. Ivanko, M.L., Domack, M.S., Stoner, M.C. & Hehir, A.R. (2016). Cost-Benefit Analysis for the Advanced Near Net Shape Technology (ANNST) Method for Fabricating Stiffened Cylinders. Available at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160011472.pdf>.
8. Kellogg, R., Bitten, R. & Mahr, E. (2016). An independent cost and schedule estimate process for NASA science projects. ... 2016 IEEE Aerospace

- ... Available at: <https://ieeexplore.ieee.org/abstract/document/7500904>.
9. Young, L., Josserand, T.M. & Chamberlin, E.P. (2017). Analyzing Cost Estimating and Analysis Organizations with in the United States Government. Available at: <https://www.osti.gov/servlets/purl/1477925>.
 10. Hamaker, J.W. (1994). But what will it cost? The history of NASA cost estimating. Available at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19950014325.pdf>.
 11. Zapata, E. (2017). An Assessment of Cost Improvements in the NASA COTS-CRS Program and Implications for Future NASA Missions. Available at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20170008895.pdf>.
 12. Jones, H.W. Estimating the lifecycle cost of spacesystems. Available at: https://ttu-ir.tdl.org/ttu-ir/bitstream/handle/2346/64311/ICES_2015_submission_41.pdf?sequence=1.
 13. Apgar, Henry. The Legacy of Parametric Estimating. Available at: <https://www.iceaaonline.com/ready/wp-content/uploads/2019/06/CV01-Paper-The-Legacy-of-Parametric-Estimating-Apgar.pdf>.
 14. Fisher, Gene. Cost Considerations in Systems Analysis. Available at: <https://apps.dtic.mil/dtic/tr/fulltext/u2/728481.pdf>.
 15. Hueber, C., Horejsi, K. & Schledjewski, R. (2016). Review of costestimation: methods and models for aerospace composite manufacturing. Manufacturing: Polymer & ... Available at: <https://www.tandfonline.com/doi/pdf/10.1080/20550340.2016.1154642>.
 16. Hunt, C.D. & Vanpelt, M.O. (2004). Comparing NASA and ESA cost estimating methods for humanmissions to Mars. Available at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20040075697.pdf>.
 17. Чепков І.Б., Зубарев В.В., Борохвостов В.К. Теорія озброєння. Науково-технічні проблеми та завдання. Т. 6. Воєнно-економічний аналіз життєвого циклу озброєння та військової техніки: теоретико-методологічні засади: монографія. Київ: ВД Дмитра Бураго. 2018. 475 с.
 18. NASA's cost estimation guide. 4th iss. United States. Available at: https://www.nasa.gov/sites/default/files/files/01_CEH_Main_Body_02_27_15.pdf.
 19. Четыркин Е.М. Статистические методы прогнозирования. М.: Статистика. 1977. 200 с.
 20. NATO“ [Analysis of LCC assessment of weapons and military equipment in NATO member states]. “Zb. nauk. pr. TSNDI OVT ZS Ukrainy” Coll. of scien. works of the CSRI of armaments and military equipment of AFU. K. No 3(78). Pp. 224—236.
 21. Hayhurst, M.R. & Wood, B.W. (2017). Mission Operations Cost Estimation Tool (MOCET). 2017 IEEE. Available at: <https://ieeexplore.ieee.org/abstract/document/7943686>.
 22. Friz, P.D., Klovstad, J.J., Leser, B.B. & Towle, B.C. (2018). Blind Study Validating Parametric Costing Tools PRICE True Planning and SEER-H for NASA Science Missions. AIAA SPACE. Available at: <https://arc.aiaa.org/doi/abs/10.2514/6.2018-5178>.
 23. Ivanco, M.L., Domack, M.S., Stoner, M.C. & Hehir, A.R. (2016). Cost-Benefit Analysis for the Advanced Near Net Shape Technology (ANNST) Method for Fabricating Stiffened Cylinders. Available at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160011472.pdf>.
 24. Kellogg, R., Bitten, R. & Mahr, E. (2016). An independent cost and schedule estimate process for NASA science projects. ... 2016 IEEE Aerospace ... Available at: <https://ieeexplore.ieee.org/abstract/document/7500904>.
 25. Young, L., Josserand, T.M. & Chamberlin, E.P. (2017). Analyzing Cost Estimating and Analysis Organizations with in the United States Government. Available at: <https://www.osti.gov/servlets/purl/1477925>.
 26. Hamaker, J.W. (1994). But what will it cost? The history of NASA cost estimating. Available at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19950014325.pdf>.
 27. Zapata, E. (2017). An Assessment of Cost Improvements in the NASA COTS-CRS Program and Implications for Future NASA Missions. Available at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20170008895.pdf>.
 28. Jones, H.W. Estimating the lifecycle cost of spacesystems. Available at: https://ttu-ir.tdl.org/ttu-ir/bitstream/handle/2346/64311/ICES_2015_submission_41.pdf?sequence=1.
 29. Apgar, Henry. The Legacy of Parametric Estimating. Available at: <https://www.iceaaonline.com/ready/wp-content/uploads/2019/06/CV01-Paper-The-Legacy-of-Parametric-Estimating-Apgar.pdf>.
 30. Fisher, Gene. Cost Considerations in Systems Analysis. Available at: <https://apps.dtic.mil/dtic/tr/fulltext/u2/728481.pdf>.
 31. Hueber, C., Horejsi, K. & Schledjewski, R. (2016). Review of costestimation: methods and models for aerospace composite manufacturing. Manufacturing: Polymer & ... Available at: <https://www.tandfonline.com/doi/pdf/10.1080/20550340.2016.1154642>.
 32. Hunt, C.D. & Vanpelt, M.O. (2004). Comparing NASA and ESA cost estimating methods for humanmissions to Mars. Available at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20040075697.pdf>.
 33. Chepkov, I.B., Zubarev, V.V. & Borokhvostov, V.K. (2018). “Teoriia ozbroennia. Naukovo-tehnichni problemy ta zavdannia. T. 6. Voенно-економічний аналіз життєвого циклу озброєння та військової техніки: теоретико-методологічні засади: монографія. Київ: ВД Дмитра Бураго. 2018. 475 с.

REFERENCES

1. “Pro oboronni zakupivli” [On Defense Procurement]. Law of Ukraine of July 17, 2020. No 808-IX. Bull. of Verkhovna Rada of Ukraine. 2020. No 67. St. 223.
2. NATO Publication ALCCP-01. NATO GUIDANCE ON LIFE CYCLE COSTS.
3. Korobchenko, S.O. (2019). “Protsey otsinky vytrat za proektamy v NASA” [NASA projects cost estimating process], “Ozbroyennia i viyskova tekhnika” [Weapons and military equipment]. K. No 4. Pp. 3—11.
4. Korobchenko, S.O. & Shapoval, P.I. (2020). “Analiz systemy otsinky vartosti zhyttyevoho tsykladu zrazkiv ozbroennia ta viyskovoi tekhniki v krainakh – chlenakh

- tekhniky: teoretyko-metodolohichni zasady: monografiia“ [The theory of armament. Scientific and technical problems and tasks. Vol. 6. Military-economic analysis of the life cycle of weapons and military equipment: theoretical and methodological principles: monograph], VD Dmitry Burago, K. 475 p.
18. NASA's cost estimation guide. 4th iss. United States. Available at: https://www.nasa.gov/sites/default/files/files/01_CEH_Main_Body_02_27_15.pdf.
19. Chetyrkyn, E.M. (1977). “Statisticheskie metody prognozirovaniia” [Statistics methods of forecasting]. Statistics, M. 200 p.

Oliynik I., Korobchenko S., Shostak V.

METHODOLOGICAL BASIS FOR PARATYMIC ASSESSMENT OF THE ARMAMENTS LIFE CYCLE COST

The article considers the methodological features of the application of regression analysis for parametric estimation of cost indicators of weapons using statistical relationships between historical data on costs and characteristics of complex technical systems. The use of life cycle cost calculations should support the process at each stage of the program, through which public customers have the opportunity to make optimal decisions according to the options presented to them. These options may include estimating future costs, comparing alternative solutions, managing existing budgets and assessing cost-cutting opportunities. The life cycle cost calculations are also used to assess the completeness of the data and determine the cost factors associated with key performance indicators of the system or key consumer requirements. The process of parametric cost estimation is analyzed, which consists of determining the main estimation hypotheses, data collection on potential cost drivers, data estimation and normalization, regression analysis, testing and selection of acceptable models. The essence of the concepts of complete, unexplained and explained deviation is illustrated and their mathematical expressions are given in the context of use in the calculations of the coefficient of determination, standard error of estimation, as well as the confidence interval and the prediction interval. The main limitations and risks of misinterpretation of the obtained results, with misinterpretation of the concepts of statistical dependence and causation have been studied. A sequence of measures for selecting the most acceptable model for its further use in forecasting is proposed.

Keywords: life cycle cost, parametric cost assessment, regression analysis, multicollinearity, armaments and military equipment, NASA, least squares method, forecasting interval, adjusted deterministic coefficient, multiple regression.

Відомості про авторів:

Олійник Іван Іванович

доктор технічних наук, старший науковий співробітник, провідний науковий співробітник Центрального науково-дослідного інституту озброєння та військової техніки Збройних Сил України, м. Київ, Україна

Information about the authors:

Ivan Oliynik

Doctor of Technical Sciences, Senior Research Officer, Lead researcher of the Central Scientific Research Institute of Armament and Military Equipment of Armed Forces of Ukraine, Kyiv, Ukraine

Коробченко Сергій Олегович

провідний науковий співробітник Центрального науково-дослідного інституту озброєння та військової техніки Збройних Сил України, м. Київ, Україна
<https://orcid.org/0000-0001-7650-5935>
email: s.korobchenko@mil.gov.ua

Serhii Korobchenko

Lead researcher of the Central Scientific Research Institute of Armament and Military Equipment of Armed Forces of Ukraine, Kyiv, Ukraine
<https://orcid.org/0000-0001-7650-5935>
email: s.korobchenko@mil.gov.ua

Шостак Владислав Григорович

в. о. директора Департаменту військово-технічної політики МОУ, м. Київ, Україна

Vladyslav Shostak

acting director of Military-technical policy department MDU, Kyiv, Ukraine

Стаття надійшла до редколегії 04.02.2021.